- SUBSTITUTE PAGE -

NOVEL MUTATED MAMMALIAN CELLS AND ANIMALS

The present application claims priority to U.S. Application Number 09/728,445 which was filed November 30, 2000 which claimed priority to U.S. Provisional Application Number 60/168,358 which was filed December 1, 1999. The present application incorporates U.S. Patent No. 6,080,576 and U.S. Applications Ser. Nos. 08/726,867, 08/728,963, 08/907,598, 08/942,806, 60/109,302, and 09/276,533 and their respective disclosures herein by reference in their entirety.

1.0. FIELD OF THE INVENTION

The present invention is in the field of molecular genetics. The application discloses novel mutated cells that are generated by process involving the insertion of at least a portion of a genetically engineered viral vector into the chromosome. The specifically disclosed recombinant vector allows for the rapid identification of the gene that has been mutated by using nucleotide or amino acid sequence information to identify the gene that has been mutated by the vector. When mutated embryonic stem cell clones are produced, such cells can be used to produce mutant animals capable of germline transmission of the described mutated genes.

2.0. BACKGROUND OF THE INVENTION

Most mammalian genes are divided into exons and introns. Exons are the portions of the gene that are spliced into mRNA and encode the protein product of a gene. In genomic DNA, these coding exons are often divided by noncoding intron sequences. Although RNA polymerase

transcribes both intron and exon sequences, the intron sequences must be removed from the transcript so that the resulting mRNA can be translated into protein. Accordingly, all mammalian, and most eukaryotic, cells have the machinery to splice exons to produce mRNA. Gene trap vectors have been designed to insert into the introns of genes in a manner that allows the cellular splicing machinery to splice vector encoded exons to cellular mRNAs. Commonly, gene trap vectors contain selectable marker sequences that are preceded by strong splice acceptor sequences and are not preceded by a promoter. Thus, when such vectors integrate into a gene, the cellular splicing machinery splices exons from the trapped gene onto the 5' end of the selectable marker sequence. Typically, such selectable marker genes can only be expressed if the vector encoding the gene has integrated into an intron. The resulting gene trap events are subsequently identified by selecting for cells that can survive selective culture.

Gene trapping has generally proven to be an efficient method of mutating large numbers of genes. The insertion of the gene trap vector creates a mutation in the trapped gene, and also provides a molecular tag for ease of identifying the gene that has been trapped. When ROSAβgeo was used to trap genes it was demonstrated that at least 50% of the resulting mutations resulted in a phenotype when examined in mice. This indicates that the gene trap insertion vectors are useful mutagens. Although a powerful tool for mutating genes, the potential of the method has historically been limited by the difficulty in identifying the trapped genes. Methods that have been used to identify trap events rely on the fusion transcripts resulting from

the splicing of exon sequences from the trapped gene to sequences encoded by the gene trap vector. Common gene identification protocols used to obtain sequences from these fusion transcripts include 5' RACE, cDNA cloning, and cloning of genomic DNA surrounding the site of vector integration. However, these methods have proven labor intensive, not readily amenable to automation, and generally impractical for high-throughput.

More recently, vectors have been developed that rely on a new strategy of gene trapping that uses a vector that contains a selectable marker gene preceded by a promoter and followed by a splice donor sequence instead of a polyadenylation sequence. These vectors do not provide selection unless they integrate into a gene and subsequently trap downstream exons which provide a polyadenylation sequence. Integration of such vectors into the chromosome results in the splicing of the selectable marker gene to 3' exons of the trapped gene. These vectors provide a number of advantages. They can be used to trap genes regardless of whether the genes are normally expressed in the cell type in which the vector has integrated. In addition, cells harboring such vectors can be screened using automated (*e.g.*, 96-well plate format) gene identification assays such as 3' RACE (see generally, Frohman, 1994, PCR Methods and Applications, *4*:S40-S58). Using these vectors it is possible to produce large numbers of mutations and rapidly identify the mutated, or trapped, gene by DNA sequence analysis.

3.0. SUMMARY OF THE INVENTION

The subject invention provides numerous isolated mammalian mutant cell clones that are each characterized by the insertion of a mutagenic genetically engineered polynucleotide sequence into a gene identifiable as corresponding to one or more of the OMNIBANK gene trapped sequences (GTSs) disclosed in Sequence Listing.

The subject invention further contemplates a mutated cell, and particularly a mutated ES cell, and the animals derived from such ES cell that stably maintain a genetically engineered mutation in a gene identifiable as corresponding to one of the disclosed GTSs.

4.0. DESCRIPTION OF THE SEQUENCE LISTING AND FIGURES

The Sequence Listing is a compilation of nucleotide sequences obtained by sequencing clonal lines of gene trapped murine ES cells.

Figures 1A-1C present a diagrammatic representation of representative gene trap vectors used to generate the described sequences.

Figure 2 provides an index to the Sequence Listing and the corresponding database accession numbers for the genes that have been mutated according to the present invention.

5.0. DETAILED DESCRIPTION OF THE INVENTION

The current invention relates to novel mutated mammalian cells that are each characterized by the insertion of a recombinant (*i.e.*, genetically engineered) mutagenic polynucleotide sequence into a gene identifiable as corresponding to one of the GTSs of SEQ ID NOS: 1-891.

For the purposes of the present invention, the term "identifiable" is to be construed as indicating that a mammalian cell, and preferably, a murine ES cell, has been mutated by the insertion of a polynucleotide sequence of recombinantly manipulated origin at a genetic locus that normally comprises polynucleotide sequence, and/or post-spliced exonic sequence, that is at least partially described in one of the GTSs of Sequence Listing. One method of determining whether one of the described mutated mammalian cells has a mutation in a gene of interest is by comparing the polynucleotide sequence (or a corresponding amino acid sequence) of the GTS identifying the mutated locus to the full length sequence of the gene. Alternatively, such searches can be conducted by comparing the described GTS sequence to a well known database (such as, but not limited to GENBANK) using established computer algorithms including, but not limited to, BLASTX, FASTA, BLASTN, BLASTP, TBLASTN, and TBLASTX using the default parameters used, for example, at the National Center for Biotechnology Information web site (www.ncbi.nlm.nih.gov). The GTSs reported in the Sequence Listing have been compared to such a database (GENBANK), and the accession numbers of the genes that have been mutated are presented in Figure 2. Accordingly, an additional aspect of the subject invention includes mutated mammalian, preferably murine, cells, or isolated cell lines, that have at least one engineered mutation in a gene identified by GENBANK or GENESEQ (for example) accession number in Figure 2.

As used herein, the terms "mutated" or "mutation" mean that the genetic locus has been altered by a process involving the integration or incorporation of a genetically

engineered polynucleotide sequence into the genome of the cell with the result that the subsequent levels of activity of the product normally encoded by the locus is altered (*i.e.*, reduced, increased, or substantially ablated). In those instances where the mutation substantially completely disrupts the expression or activity of the product normally encoded by the locus (*i.e.*, a null mutation), a cell that is heterozygous for the mutated allele will typically produce about one half of the product of a nonmutated cell (via a gene dosage effect), and about twice the amount of product produced by a cell that is homozygous for the mutant allele.

The term "recombinantly manipulated" shall mean that such compositions comprising such molecules or polynucleotides have been genetically engineered using molecular biology methodologies *in vitro* or *ex vivo* (see generally, Sambrook *et al.*, 1989, Molecular Cloning, A Laboratory Manual, Cold Springs Harbor Press, N.Y.; and Ausubel *et al.*, 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y.).

Where, the specifically exemplified mammalian cells, *i.e.*, embryonic stem cells (Lex-1 cells from murine strain A129), are mutated by a process involving the insertion of at least a portion of a genetically engineered vector sequence into the gene of interest, the mutated embryonic stem cells can be microinjected into blastocysts which are subsequently introduced into pseudopregnant female hosts and carried to term using established methods such as those described in, for example, "Mouse Mutagenesis", 1998, Zambrowicz *et al.*, eds., Lexicon Press, The Woodlands, TX,

and periodic updates thereof, herein incorporated by reference. The resulting chimeric animals are subsequently bred to produce offspring capable of germline transmission of an allele containing the engineered mutation in the gene of interest.

An alternative method of producing mutated cells and animals in the specifically exemplified genes involves the process of gene targeting by homologous recombination using methods such as those exemplified in U.S. Application Ser. No. 09/171,642, which is herein incorporated by reference in its entirety. Mutations produced using such methods include, but are not limited to knockout mutations, "knockin" mutations (where a human gene, for example, is used to replace its murine orthologs), can be conditional, can include point mutations, and mutations that activate gene expression. Some of the mutations described above (conditional mutations, point mutations, etc.) can be produced via processes that involve the substantial removal of vector encoded sequences (often recombines mediated) subsequent to the incorporation of the recombinantly manipulated sequences into the genome.

5.1. MUTATED MAMMALIAN CELLS OF THE PRESENT INVENTION

The presently described mutated cells have genetically engineered mutations in genes identifiable as corresponding to, or normally comprising, at least a portion of a sequence reported in the Sequence Listing as SEQ ID NOS: 1-891. Additional embodiments of the present invention are cells comprising engineered mutations in homologs, paralogs, orthologs, etc., of the mutated genes disclosed in the Sequence Listing. Such homologs,

paralogs, and orthologs include genes having sequences that hybridize to one or more of the disclosed GTSs of SEQ ID NOS: 1-891 under stringent, or preferably highly stringent, conditions. Hybridization conditions also provide an alternative means of identifying the mutated genes corresponding to the GTSs reported in the sequence listing. Typically, such genes will be identifiable because a disclosed GTS, or portion thereof, shall hybridize to the gene under stringent conditions.

By way of example and not limitation, high stringency hybridization conditions can be defined as follows: Prehybridization of filters containing DNA to be screened is carried out for 8 h to overnight at 65°C in a buffer containing 6X SSC, 50mM Tris-HCl (pH 7.5), 1mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 μg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C in prehybridization mixture containing 100μg/ml denatured salmon sperm DNA and 5-20 x 10^6 cpm of ^{32}P-labeled probe (alternatively, as in all hybridizations described herein, approximately 42, 44, 46, 48, 50, 52, 54, 56, 58, 62, 64, 66, 68, 70, or about 72 degrees or more can be used). The filters are then washed in approximately 1X wash mix (10X wash mix contains 3M NaCl, 0.6M Tris base, and 0.02M EDTA, alternatively, as with all washes described herein, 2X, 3X, 4X, 5X, 6X wash mix, or more, can be used) twice for 5 minutes each at room temperature, then in 1X wash mix containing 1% SDS at 60°C (alternatively, as in all washes described herein, approximately 42, 44, 46, 48, 50, 52, 54, 56, 58, 62, 64, 66, 68, 70, or about 72 degrees or more can be used) for about 30 min, and finally in 0.3X wash mix (alternatively, as in all final washes described herein,

approximately, 0.2X, 0.4X, 0.6X, 0.8X, 1X, or any concentration between about 2X and about 6X can be used in conjunction with a suitable wash temperature) containing 0.1% SDS at 60°C (alternatively, approximately 42, 44, 46, 48, 50, 52, 54, 56, 58, 62, 64, 66, 68, 70, or about 72 degrees or more can be used) for about 30 min. The filters are then air dried and exposed to x-ray film for autoradiography. In an alternative protocol, washing of filters is done for 37°C for 1 h in a solution containing 2X SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA. This is followed by a wash in 0.1X SSC at 50°C for 45 min before autoradiography. Another example of hybridization under highly stringent conditions is hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1xSSC/0.1% SDS at 68°C (Ausubel F.M. *et al.*, eds., 1989, Current Protocols in Molecular Biology, Vol. I, Green Publishing Associates, Inc., and John Wiley & sons, Inc., New York, at p. 2.10.3). Alternatively, moderately stringent conditions can be used (*e.g.*, washing in 0.2xSSC/0.1% SDS at 42° C (Ausubel *et al.*, 1989, *supra*). Moderately stringent conditions can be additionally defined, for example, as follows: Filters containing DNA are pretreated for 6 h at 55°C in a solution containing 6X SSC, 5X Denhart's solution, 0.5% SDS and 100 µg/ml denatured salmon sperm DNA. Hybridizations are carried out in the same solution and 5-20 x 10^6 cpm ^{32}P-labeled probe is used. Filters are incubated in hybridization mixture for 18-20 h at 55°C (alternatively, as in all hybridizations described herein, approximately 42, 44, 46, 48, 50, 52, 54, 56, 58, 62, 64, 66, 68, 70, or about 72 degrees or more can be used in combination with a

suitable concentration of salt). The filters are then washed in approximately 1X wash mix (10X wash mix contains 3M NaCl, 0.6M Tris base, and 0.02M EDTA, alternatively, as with all washes described herein, 2X, 3X, 4X, 5X, 6X wash mix, or more, can be used) twice for 5 minutes each at room temperature, then in 1X wash mix containing 1% SDS at 60°C (alternatively, as in all washes described herein, approximately, 42, 44, 46, 48, 50, 52, 54, 56, 58, 62, 64, 66, 68, 70, or about 72 degrees or more can be used) for about 30 min, and finally in 0.3X wash mix (alternatively, as in all final washes described herein approximately 0.2X, 0.4X, 0.6X, 0.8X, 1X, or any concentration between about 2X and about 6X can be used in conjunction with a suitable wash temperature) containing 0.1% SDS at 60°C (alternatively, approximately 42, 44, 45, 48, 50, 52, 54, 56, 58, 62, 64, 66, 68, 70, or about 72 degrees or more can be used) for about 30 min. The filters are then air dried and exposed to x-ray film for autoradiography.

In an alternative protocol, washing of filters is done twice for 30 minutes at 60°C in a solution containing 1X SSC and 0.1% SDS. Filters are blotted dry and exposed for autoradiography.

Other conditions of moderate stringency which may be used are well-known in the art. For example, washing of filters can be done at 37°C for 1 h in a solution containing 2X SSC, 0.1% SDS. Another example of hybridization under moderately stringent conditions is washing in 0.2xSSC/0.1% SDS at 42°C (Ausubel et al., 1989, *supra*). Such less stringent conditions may also be, for example, low stringency hybridization conditions. By way of example and not limitation, procedures using such

conditions of low stringency are as follows (see also Shilo and Weinberg, 1981, Proc. Natl. Acad. Sci. USA 78:6789-6792): Filters containing DNA are pretreated for 6 h at 40°C in a solution containing 35% formamide, 5X SSC, 50mM Tris-HCl (pH 7.5), 5mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 μg/ml denatured salmon sperm DNA. Hybridizations are carried out in the same solution with the following modifications: 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100μg/ml salmon sperm DNA, 10% (wt/vol) dextran sulfate, and 5-20 X 10^6 cpm ^{32}P-labeled probe is used. Filters are incubated in hybridization mixture for 18-20 h at 40°C (alternatively, as in all hybridizations described herein, approximately 42, 44, 46, 48, 50, 52, 54, 56, 58, 62, 64, 66, 68, 70, or about 72 degrees or more can be used). The filters are then washed in approximately 1X wash mix (10x wash mix contains 3M NaCl, 0.6M Tris base, and 0.02M EDTA, alternatively, as with all washes described herein, 2X, 3X, 4X, 5X, 6X wash mix, or more, can be used) twice for five minutes each at room temperature, then in 1X wash mix containing 1% SDS at 60°C (alternatively, as in all washes described herein, approximately 42, 44, 46, 48, 50, 52, 54, 56, 58, 62, 64, 66, 68, 70, or about 72 degrees or more can be used) for about 30 min, and finally in 0.3X wash mix (alternatively, as in all final washes described herein, approximately, 0.2X, 0.4X, 0.6X, 0.8X, 1X, or any concentration between about 2X and about 6X can be used in conjunction with a suitable wash temperature) containing 0.1% SDS at 60°C (alternatively, approximately 42, 44, 46, 48, 50, 52, 54, 56, 58, 62, 64, 66, 68, 70, or about 72 degrees or more can be used) for about 30 min. The filters are then air dried and exposed to x-ray film for

autoradiography. In yet another alternative protocol, washing of filters is done for 1.5 h at 55°C in a solution containing 2X SSC, 25mM Tris-HCl (pH 7.4), 5mM EDTA, and 0.1% SDS. The wash solution is replaced with fresh solution and incubated an additional 1.5 h at 60°C. Filters are then blotted dry and exposed for autoradiography. If necessary, filters are washed for a third time at 65-68°C and reexposed to film. Other conditions of low stringency which may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). Preferably, GTS variants identified or isolated using the above methods will also encode a functionally equivalent gene product (*i.e.*, protein, polypeptide, or domain thereof, encoding or otherwise associated with a function or structure at least partially encoded by the complementary GTS).

Low stringency conditions are well known to those of skill in the art, and will vary predictably depending on the specific organisms from which the library and the labeled sequences are derived. For guidance regarding such conditions see, for example, Sambrook *et al.*, 1989, Molecular Cloning, A Laboratory Manual, Cold Springs Harbor Press, N.Y.; and Ausubel *et al.*, 1989, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y.

The identification of homologs, heterologs, or paralogs of SEQ ID NOS: 1-891 in other, preferably related, species can be useful for developing additional animal model systems that are closely related to humans for purposes of drug discovery. Genes at other genetic loci within the genome that encode proteins which have extensive

homology to one or more domains of the gene products encoded by SEQ ID NOS: 1-891 can also be identified via similar techniques. In the case of cDNA libraries, such screening techniques can identify clones derived from alternatively spliced transcripts in the same or different species.

Techniques useful to disrupt a gene in a cell and especially an ES cell that may already have a disrupted gene are disclosed in copending US patent applications Nos. 08/726,867; 08/728,963; 08/907,598; and 08/942,806, all of which are hereby incorporated herein by reference in their entirety, are within the scope of the current invention to disrupt a gene that encodes a polynucleotide of the current invention.

5.2. USES OF THE DESCRIBED MUTATED GENES AND ANIMALS

The described mutated cells and animals are used to investigate and define the cellular and biological functions of the mutated genes. Producing a scientific model that accurately accounts for the large number of genes, proteins, and macromolecules within a single cell has thus far proved beyond the capabilities of existing computer technology. It should thus not be surprising that the far more complex task of modeling the various intricacies, cross and direct redundancies, and interrelated functions of the various metabolic and catabolic processes that occur within a single cell has also proven largely intractable to algorithmic methods of modeling and prediction. Even if one assumes that computer modeling of inherently chaotic/heuristic processes will rapidly mature in the near future, such methods, at best,

can only provide predictions that subsequently require practical validation. Several decades of empirical data have proven that mutant phenotypes provide a valuable source of such validation. The mutated diploid mammalian cells of the present invention will initially exist as mutated diploid cells that are heterozygous (except where genes on the X or Y chromosomes are mutated) for the mutations identified in the sequence listing. As such, via a "gene dosage" effect, the mutated cells can typically be characterized by the fact that they produce about one half of the mutated transcript/activity relative to cells having two nonmutated or wild type copies of the corresponding gene.

When mutant animals are produced from the mutated cells, heterozygous animals capable of germline transmission of the mutated allele can be bred to produce embryos or offspring that are homozygous for the mutant allele. Such animals or embryos are a rich source of tissues and cells that do not express physiologically relevant amounts of the mutated genes or activities encoded thereby. Accordingly, an additional embodiment of the present invention are mutant cells and animals that have homozygous mutations in genes identifiable as corresponding to the GENBANK, or other database accession, numbers provided in Figure 2, or are identifiable as a homologs, paralog, or orthologs of a sequence provided in the Sequence Listing.

In addition to providing important information regarding the functional role of a given gene in its nonmutated state (*i.e.*, you learn about the function of the gene by discerning the effects of reducing or ablating the

activity normally encoded by the gene), the described mutated cells and animals can be used as disease models, or in assays for compounds or genes (via gene delivery or transgenic methods) that compensate for the mutant phenotype and that can be used to treat diseases and disorders related to the observed phenotype. Alternatively, such products and genes can also be used to enhance desirable, if not normal, symptoms related to the observed phenotypes.

The gene replacement/delivery therapies described above should be capable of delivering gene sequences to the cell types within patients which express the peptide or protein having the desired activity.

The examples below are provided to illustrate the subject invention. These examples are provided by way of illustration and are not included for the purpose of limiting the invention in any way whatsoever.

6.0. EXAMPLES

6.1. GENERATION OF A LIBRARY OF MUTATED MOUSE ES CELLS DEFINED BY GTS SEQUENCES

The retroviral vector VICTR 3, described in detail in U.S. application Ser. No. 08/728,963, filed October 11, 1996, was used to generate a library of gene trapped ES cell clones that represent a portion of the described GTSs. A plasmid containing the VICTR 3 cassette was constructed by conventional cloning techniques and designed to employ the features described above. Namely, the cassette contained a *PGK* promoter directing transcription of an exon that encodes the *puro* marker and ends in a canonical splice donor sequence. At the end of the puromycin exon,

sequences were added as described that allow for the annealing of two nested PCR and sequencing primers. The vector backbone was based on pBluescript KS+ from Stratagene Corporation.

The plasmid construct was linearized by digestion with Sca I which cuts at a unique site in the plasmid backbone. The plasmid was then transfected into the mouse ES cell line AB2.2 by electroporation using a BioRad Genepulser apparatus. After the cells were allowed to recover, gene trap clones were selected by adding puromycin to the medium at a final concentration of 3 μg/ml. Positive clones were allowed to grow under selection for approximately 10 days before being removed and cultured separately for storage and to determine the sequence of the disrupted gene.

Total RNA was isolated from an aliquot of cells from each of 18 gene trap clones chosen for study. Five micrograms of this RNA was used in a first strand cDNA synthesis reaction using the "RS" primer. This primer has unique sequences (for subsequent PCR) on its 5' end and nine random nucleotides or nine T (thymidine) residues on it's 3' end. Reaction products from the first strand synthesis were added directly to a PCR with outer primers specific for the engineered sequences of puromycin and the "RS" primer. After amplification, an aliquot of reaction products were subject to a second round of amplification using primers internal, or nested, relative to the first set of PCR primers. This second amplification provided more reaction product for sequencing and also provided increased specificity for the specifically gene trapped DNA.

The products of the nested PCR were visualized by agarose gel electrophoresis, and seventeen of the eighteen clones provided at least one band that was visible on the gel with ethidium bromide staining. Most gave only a single band which is an advantage in that a single band is generally easier to sequence. The PCR products were sequenced directly after excess PCR primers and nucleotides were removed by filtration in a spin column (Centricon-100, Amicon). DNA was added directly to dye terminator sequencing reactions (purchased from ABI) using the standard M13 forward primer a region for which was built into the end of the *puro* exon in all of the PCR fragments.

Subsequent studies have used both VICTR 3 and VICTR 20. Like VICTR 3, VICTR 20 is exemplary of a family of vectors that incorporate two main functional units: a sequence acquisition component having a strong promoter element (phosphoglycerate kinase 1) active in ES cells that is fused to the puromycin resistance gene (or other exon sequence) that is followed by a synthetic consensus splice donor (SD) sequence and lacks an operatively positioned polyadenylation sequence downstream from the SD sequence (PGKpuroSD); and 2) a mutagenic component that incorporates a splice acceptor sequence fused to a selectable and/or colorimetric marker gene and followed by a polyadenylation sequence (for example, SA3geopA, SAneopA, SAIRESneopA, or SAIRESβgeopA).

Also like VICTR 3, stop codons have been engineered into all three reading frames in the region between the 3' end of the selectable marker and the splice donor site. A diagrammatic description of structure and functions of VICTRs 3 and 20 is provided in Figure 1.

When VICTRs 3, 20, and various variations thereof such as the vectors and methods described in U.S. Applications Ser. Nos. 09/276,533, and 60/095,989 (the disclosures of which are herein incorporated by reference), were used in the commercial scale application of the presently disclosed invention, many mutagenized ES cell clones were rapidly engineered and obtained. Sequence analysis obtained from these clones has identified a wide variety of sequences. Each of the sequences presented in SEQ ID NOS: 1-891 identify novel mutations in the coding regions of mammalian genes that identifiable as corresponding to the sequences presented in the Sequence Listing. Alternatively, the described mutated cells are described by the database (GENBANK, GENSEQ, etc.) accession numbers for the corresponding genes that have been mutated (see Figure 2). The described mutated cells, and preferably ES cells, provide a valuable resource for defining, evaluating, or validating the biological function or disease/pharmaceutical relevance of each of these genes.

The cloned 3' RACE products resulting after the target ES cells were infected with one of the described gene trap vectors were purified using conventional column chromatography, (*e.g.*, S300 and G-50 columns), and the products were recovered by centrifugation. Purified PCR products were quantified by fluorescence using PicoGreen (Molecular Probes, Inc., Eugene Oregon) as per the manufacturer's instructions.

Dye terminator cycle sequencing reactions with AmpliTaq® FS DNA polymerase (Perkin Elmer Applied Biosystems, Foster City, CA) were carried out using approximately 7 pmoles of sequencing primer, and

approximately 30-120 ng of 3' template. Unincorporated dye terminators were removed from the completed sequencing reactions using G-50 columns as described above. The reactions were dried under vacuum, resuspended in loading buffer, and electrophoresed through a 6% Long Ranger acrylamide gel (FMC BioProducts, Rockland, ME) on an ABI Prism® 377 with XL upgrade as per the manufacturer's instructions. The sequences of the resulting amplicons, or GTSs, are described in SEQ ID NOS: 1-891. All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.